DOCUMENT RESUME

ED 222 573 TM 820 738

AUTHOR Weber, George
TITLE The National Assessment of Educational Progress.
PUB DATE 2 Aug 82
NOTE 23p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Achievement Tests; *Criterion Referenced Tests; Educational Assessment; Educational Objectives; *Elementary School Students; Evaluation Criteria; *Federal Programs; National Programs; *Program Evaluation; *Secondary School Students
IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

The National Assessment of Educational Progress (NAEP) is a federally funded, nationwide survey of educational achievement in 10 academic fields. The program is examined to determine whether viable alternatives exist to measure educational achievement. The information produced by NAEP concerns the educational achievement of 9-year-olds; 13-year-olds; 17-year-olds; and young adults, ages 26 to 35. The program is helping to develop a consensus on educational objectives and is making important contributions to the ability to measure educational achievement. The ultimate purpose of NAEP is to improve the educational achievement of school age children. The major achievements of the program thus far include a list of educational objectives, a large body of reliable national data on educational achievement in terms of these objectives, and several contributions to the improvement of the methodology of measuring educational achievement. The achievement testing programs that have been established since NAEP was begun do not provide comparable data. NAEP data are unique because they are national, cover 10 academic fields, and are domain-referenced. The author believes the program should be expanded to realize its potential. Recommendations for expansion and improvement are discussed. (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED222573

August 2, 1982
3505 Halcyon Drive
Alexandria, Va. 22305
(703) 549-5284

THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

by

George Weber

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Summary

The National Assessment of Educational Progress (NAEP) is providing a large body of unique, valuable information about the educational achievement of our children and young people. It is also helping to develop a consensus on educational objectives and making important contributions to our over-all ability to measure educational achievement. Because there are no reasonable alternatives at this time for producing similar results (e.g., by using data collected as part of other testing programs), NAEP should be continued. In fact, expansion of the project along certain lines would be desirable. The project, for various reasons, is not realizing its potential for raising the educational achievement of our school-age children and youth. Several recommendations are offered for its improvement.

TM 820 738

Contents

Introduction. 3

What NAEP Is. 3

What NAEP Aims to Do. 4

How NAEP Tries to Achieve this Purpose. 6

What NAEP Does Not Do 6

The Achievements of NAEP to Date. 7

Alternatives to NAEP. 10

The Cost of NAEP. 15

Use of NAEP Data. 16

Recommendations for Expansion of NAEP 20

Other Recommendations 22

Introduction

This paper addresses the specific question: Are there reasonable alternatives to the National Assessment of Educational Progress (NAEP); e.g., using data collected as part of other testing programs? In order to do this meaningfully, a broad examination of NAEP must be made, as well as an examination of other testing programs. In making such an examination, I have developed some recommendations for the improvement of NAEP, which I believe should be continued and in certain ways expanded.

What NAEP Is

NAEP is a federally funded, nationwide survey of educational achievement in ten academic fields: reading, writing, mathematics, literature, science, social studies, music, art, citizenship, and career development. The survey covers four age groups: 9-year-olds, 13-year-olds, 17-year-olds, and young adults (26-35). (In recent years, most assessments have not included out-of-school 17-year-olds or young adults.) NAEP uses a stratified sampling method and produces data for four regions of the country, the two sexes, race and ethnic group (three categories), level of parental education (four categories), and type of community (seven categories). Each testee takes only a small part of the complete survey. The original plan was to survey each of the ten fields every five years on a rotating basis, but the schedule has been cut back so that now only reading, writing, and mathematics are on a five-year cycle; the other seven fields are being assessed only as funds allow. The knowledge, skills, understandings, and attitudes assessed are those deemed important for all students by representatives of three groups: scholars, educators, and laymen. The particular exercises in the tests aim at determining what almost all (90%) of the age-group can do, what about half can do, and what only a few (10%) can do.

About one-third of the items are in each of these categories. Some but not all of the items use the multiple-choice format. Some of the items are released after each assessment; the rest are kept secret for later use to provide comparisons over time. NAEP has always claimed that it passes no judgments on the results; in other words, that it produces factual data without commenting on whether these data are cause for satisfaction or alarm. (In reality, this is sometimes not the case, and comment on this matter appears later in this paper.) NAEP was first proposed in the summer of 1963, and the first survey (in science, citizenship, and writing) began in the spring of 1969. We have now had surveys in all ten fields and as many as three in some.

What NAEP Aims to Do

The ultimate purpose of NAEP is to raise the educational achievement of school-age children and youth. It is not always stated so baldly. It was often said when NAEP was being developed (1963-68) that the purpose was to compile censuslike data about educational achievement, data that was then not available on a nationwide basis. In fact, little achievement data of any kind was available, although there was much information about enrollment, expenditures, number of schools, length of schooling, and so forth -- all "input" data. It was as if we knew how many cattle farms we had, how much feed was being used, what diseases vets treated, but not how much beef went to market. While this was true, and Francis Keppel, then U.S. Commissioner of Education, could correctly point out that he had little solid evidence with which to carry out his responsibility to report to Congress on the condition of American education, it is hard to believe that NAEP would have gotten very far if its sole objective had been to produce disinterested,

censuslike data.

The motivating force behind NAEP was the widespread belief, indeed conviction, in the early 1960's that educational achievement was seriously deficient. Arthur Bestor's EDUCATIONAL WASTELANDS (1953) was only one of the academic criticisms of the educational status quo during the previous decade. On the popular side, Rudolf Flesch's WHY JOHNNY CAN'T READ (1955) was a best-seller. The Council for Basic Education, a reform group, was founded in 1956. Sputnik went up in 1957, and the National Defense Education Act was passed the following year. By the early Sixties, the formidable Admiral H. G. Rickover was suggesting national standards and national tests. The very appointment of "outsiders" like Sterling McMurrin and Francis Keppel to the position of U.S.Commissioner showed President Kennedy's concern about educational achievement at the elementary and secondary levels. And no one could have followed the delicacy with which Ralph Tyler dealt with the strong and outspoken opposition to the proposed assessment from the National Association of School Administrators without sensing that the leaders of that organization, made up of the public school superintendents, felt that the real purpose of the assessment was to prove poor educational achievement so that something would be done about it.

But to gain the grudging acceptance of the public school establishment and its essential cooperation (some superintendents said they would never let the assessors into their schools, and some did not), the NAEP had to be set up in such a way that no state or school district data could be produced. This decision limited the usefulness of NAEP in the short run, but NAEP's approach has proved so valuable that it has been a major factor leading to the establishment of statewide testing programs in some 36 states and in many local school districts. The forces demanding accountability during a

period when there was a widespread public perception of the schools' shortcomings were too strong to be denied.

In practical terms, then, what NAEP aims to do is to improve the educational achievement of school-age children and youth. To what degree it has succeeded will be discussed later.

How NAEP Tries to Achieve this Purpose

Because public schooling is fundamentally a state, not a national, function, and operationally a local one, NAEP has never tried to achieve educational improvement directly. To do so would have run up against the traditional opposition to national standards, national curricula, and Federal control of education. (This opposition is so strong that no general Federal aid bill has ever been passed, and the closest thing to it, the Elementary and Secondary Education Act of 1965, contained explicit language pretending to rule out any Federal controls over curricula and programs.) Given these taboos, NAEP rests on the belief of its creators that the facts would speak for themselves and lead, via a number of different efforts by others, to the higher educational achievement sought. What NAEP was to supply was reliable national data on educational achievement in terms of widely accepted goals of elementary and secondary schooling and, not incidentally, better means of measuring educational achievement. While these specific goals were limited, they were by no means inconsequential.

What NAEP Does Not Do

Because of the way it gathers its data, NAEP cannot evaluate the performance of a given individual, a given school, a given school district, or a given state. Nor can it evaluate specific curricula or programs. It is also limited in scope because, although it measures achievements in almost all the major academic areas, it does not cover all of them. For

example, it does not cover foreign languages or physical education. (The achievement of American students in foreign languages is abysmally low, and one wonders whether it would not be higher if the area had been included in NAEP, as it certainly would have been in a similar effort in a Western European country, the Soviet Union, or Japan. As for physical education, some reasonably reliable data are available on the basis of the achievement test developed and promoted by the President's Council on Physical Fitness and Sports.) Nor does NAEP cover the various "affective" areas that some school people believe are important (e.g., self-concept, cooperation, ambition, tolerance, morality), although some of these matters were included to a degree in the citizenship assessment and the one on career and occupational development. The founders of NAEP wanted to concentrate on academic areas, and that they did, although not without some controversy at the time.

The Achievements of NAEP to Date

The major achievements of NAEP to date are three: a list of educational objectives on which there is a wide agreement among subject-matter specialists (scholars), educators, and laymen; a large body of reliable national data on educational achievement in terms of these objectives; and a number of important contributions to the improvement of the methodology of measuring educational achievement. Each of these achievements is remarkable in view of the considerable controversy about education and the amount of opposition to testing in general and NAEP in particular. A cynic might observe, however, that these achievements were possible because NAEP was not linked to any policy questions or specific actions. That is, no teacher's job or student's diploma was in jeopardy. No educational program was directly in danger. No comparisons could be made between schools, districts, or states. Notwithstanding a degree of truth in this line of reasoning, the achievements offer the potential for considerable progress in educational achievement,

and some of the potential has already been achieved.

While the list of educational objectives is not beyond debate, its mere existence is an achievement. Here are ten academic fields, and objectives for each, on which subject-matter people, educators, and laymen have reached a consensus; a large number of diverse people have said, in effect: "These are educational achievements that all young people should aim for."

The large body of data is, when examined closely, most impressive. The following is but a tiny sample.

In reading, many young Americans are handicapped by serious reading deficiencies. Specifically, 15 to 20 percent of 17-year-olds cannot do simple functional reading tasks at an adult level. Moreover, from the first reading assessment in 1970-71 to the third one in 1979-80, there was no significant improvement in the reading achievement of 13-year-olds or 17-year-olds. Nine-year-olds, however, improved their achievement over the decade in all three reading skills: literal comprehension, inferential comprehension, and reference skills.

In writing, the general level of skills is not very high. By age 17, only about half have some mastery of the basics, but even those rarely attempt anything beyond the simplest constructions or use anything beyond a rather limited vocabulary. Neither 13-year-olds or 17-year-olds receive a great deal of direct instruction in writing or are required to do much writing in school. A considerable proportion of 17-year-olds do not understand the nature and conventions of written language.

In mathematics, about one-third of 13-year-olds cannot get the right answers to a set of four/very simple computations (addition, subtraction, mutliplication, division) with small whole numbers. That is, they missed one or more of the following: adding 38 and 19; subtracting 19 from 36;

multiplying 38 by 9; and dividing 125 by 5. About 84 percent can correctly add a column with the four decimal numbers: 3.09, 10.00, 9.14, and 5.10 — the kind of problem one might have on a sales slip. About two-thirds cannot do relatively simple word problems with easy computation.

In social studies, only 50% of 17-year-olds know the meaning of the term "monopoly". Fewer than 20% of 17-year-olds know some of the factors that determine whether a society has a low birth and death rate. Only about half know the basis for the Supreme Court decision forbidding prayer in the schools. Only about a third know that places nearer the equator are warmer than those farther away.

In music, students have only a superficial understanding of musical notation and terminology and the history of music. At all three ages (9, 13, 17), girls outshine boys in music. Eighty percent of 17-year-olds have taken a class in general music or music appreciation. One-third have participated in school vocal or instrumental groups for at least a year.

In art, most students say they enjoy art, but they are able to do so only on a simple level. They are "artistic literalists", praising a painting because it looks lifelike or because its subject is one they like. Nearly half of all 17-year-olds have never visited an art museum or have visited one only once. Few elementary schools have special art teachers. Those that do provide only 15-30 hours of specialized instruction a year. Boys are better than girls at art at age 9, but girls take the lead at 13 and 17.

In science, 75% of 17-year-olds can indicate where a weight should be placed to balance a beam in a simple physics experiment. In a multiple-choice question, 41% of 17-year-olds can correctly identify the function of the placenta in a pregnant human female. Among 13-year-olds, 48% recognize, in another multiple-choice question, that it is heat (rather than electricity, light, magnetism, or sound) that is produced by the wasted energy in driving a gasoline-powered car. At every age and in almost every area of science, male achievement is higher than female. (Incidentally, two researchers in

the teaching of science, Lee S. Shulman and Pinchas Tamir, have called NAEP "surely the largest-scale study of achievement in science.")

NAEP has made a number of important contributions to the improvement of the methodology of measuring educational achievement. Perhaps the most important is the development of criterion-referenced items rather than norm-referenced items. These are constructed in an entirely different way (an explanation is given below). In addition, by adding an "I don't know" alternative to multiple-choice questions, their validity has been improved. Further, the stratified-sampling approach has opened a whole new strategy for assessing the achievement of large groups. Finally, the development of many kinds of items outside the multiple-choice format has demonstrated the limitations of multiple-choice and provides opportunities for assessing knowledge, skills, understandings, and attitudes that previously were not being assessed. These changes have together led to a whole new direction for assessing the achievement of individuals as well as groups.

Alternatives to NAEP

Even with the new testing programs that have been established since the beginning of NAEP, in part inspired by NAEP, the data provided by other programs are not comparable. The other achievement testing programs fall into five categories:

1) course-oriented tests
2) norm-referenced tests
3) standardized criterion-referenced tests
4) school district and state minimum competency tests
5) college admission tests

All of these tests have their value for certain purposes, but none provides the kind of data that NAEP provides. Generally, it is because

none of these except norm-referenced tests provides data on a truly national basis, none has the breadth of subject coverage that NAEP does; two major categories — norm-referenced tests and college admission tests — are norm-referenced; some of the detailed results are not available to the public, and all but the first are multiple-choice in form. I will discuss each of the five kinds of tests in some detail to show why their results cannot be used as a substitute for the kind of information that we get from NAEP.

Course-oriented tests are probably the most common and most familiar achievement tests used in the schools. At the end of a course or course-segment, a test (developed typically by the teacher but sometimes by the authors of the text or the creators of a course or program) is given to see how well the students have learned what has been taught. Obviously, there are no national results on such tests. Moreover, their scope is limited. Even a broad test in an academic field, such as writing, would not cover the territory that NAEP covers. In other words, it will cover only what has been taught in a given period, not what should be learned by a certain age. Although these tests are a valuable part of instruction, they provide no data of the NAEP kind.

The norm-referenced achievement tests (e.g., Metropolitan, SRA, Iowa, Stanford, STEP) would seem capable of providing the kind of data we get from NAEP. They are supposed to be field-tested and normed on a national basis. They all cover reading and mathematics, and some cover, in addition, science, social studies, punctuation, English expression, and study skills. But this list of academic areas falls far short of the NAEP coverage. Furthermore, the detailed results are not readily available, since the test publishers want to be able to sell the tests and have them used over an extended period. (There is no reason why this could not be changed, for a price. That is,

these tests could be put out like automobiles in annual versions, with each version becoming available to the public at the end of the year. But this is not done now; the norm-referenced tests are used over many years and their exact content is kept secret.)

But the fundamental deficiencies of the data from these tests (were they available), compared to NAEP data, stems from two characteristics: they are multiple-choice and they are norm-referenced. These deficiencies are so important that they warrant some elaboration.

The multiple-choice format has a number of weaknesses, the chief of which that are pertinent to this discussion are that it involves recognition rather than recall and it requires no creation or composition. The norm-referenced tests present the student with a number of possible answers to a question, only one of which is correct. The ability to recognize the correct answer is all that is required, but in non-school life recall rather than recognition is almost always needed. For example, in "real-life" situations, one must recall that Lincoln was the Civil War president, rather than pick his name from four possibilities presented. The second weakness is also important. Although NAEP takes advantage of the economy of the multiple-choice format, it uses other formats as well. Its writing assessment, for example, asks students to write. Its science assessment, for example, asks students to work with scientific paraphernalia. If money were no object, NAEP would probably use many more exercises ~~of~~ in these other formats.

The deficiencies of the data stemming from the norm-referenced nature of these tests are even more important. Norm-referenced tests and the items that make them up are fundamentally different from criterion-referenced tests, and these differences are best understood by seeing how the two kinds are constructed.

A norm-referenced test of third-grade reading, for example, begins

with a study of what is common to the main textbooks used through third grade in present-day schools. This material is then sampled, usually by means of vocabulary and comprehension questions that differentiate between children. A question that almost all children would get right or one that almost all would get wrong is eliminated, no matter how important such a question is to reading achievement. The test results, then, will distribute the children along the familiar bell-shaped curve. That is the purpose of the test. A given child will get a score (grade-equivalent, percentile, stanine, percentile-band) that relates his achievement to the achievement of all children of the same grade placement in the nation. But except in this comparative way, it will reveal nothing about how well or how poorly he reads.

The criterion-referenced test begins differently. What, ask the test writers, should a child accomplish in reading by the end of the third grade? This question is answered in terms of specific objectives such as phonics knowledge, vocabulary, syllabication, correct tense forms, synonyms, and the meaning of common prepositions. The questions are written in terms of these objectives. One test available from a major publisher in the mid-1970's (it may still be on the market) asked third-graders 153 questions related to 41 objectives. The test results purport to show how many and which of these objectives have been mastered by the student. Such information can be compiled for a class, a school, or a school system.

The norm-referenced tests, designed to compare the performance of an individual or a group with a large (usually national) group, tell nothing directly about achievement with respect to instructional objectives. One could, if allowed to examine the most widely used norm-referenced tests, get some idea of the achievements of the average third-grader and of those who were in the top ten percent and in the bottom ten percent. But because of the way the tests are constructed, the data would not be of the NAEP type.

As Ralph Tyler once put it, "the exercises [that make up norm-referenced tests] have not been obtained by a systematic sampling of what children are expected to learn. Instead, the exercises comprise a sample of items that differentiate children. It may seem odd. . .but this is due to the fact that most psychometrists since World War I have been interested in individual differences among children and in the process of sorting children rather than in the processes of learning." (COUNCIL FOR BASIC EDUCATION BULLETIN, March 1972)

Turning to the third category of non-NAEP tests, there are now a number of standardized criterion-referenced tests in use. The way they are constructed was set forth above. Although NAEP calls its tests "domain-referenced" instead of "criterion-referenced", and there are some differences, the standardized criterion-referenced tests are similar to the NAEP tests in many respects. But they could not provide the same body of data because they are not used extensively, they do not cover the ten NAEP subject areas, and they are entirely multiple-choice in form. Accordingly, we might be able to learn from these standardized crtierion-referenced tests the achievement of some Washington, D.C., children in reading, or some New York City students in math, but we could not ascertain, even if we were able to collect the results from all these tests, the national achievement in any academic field, let alone the ten NAEP fields.

In recent years, partly as a result of the NAEP influence and assistance, some 36 states and some individual school districts have inaugurated minimum competency test programs geared to the granting of high school diplomas. These are criterion-referenced tests, and they are providing some valuable data on achievement in certain states. But all the data put together would not match the NAEP output. For one thing, the availability of the data is spotty. Further, the tests usually cover only a minimum achievement level (roughly, the ten percent level of NAEP). Moreover, the subject coverage

is almost always considerably narrower than the ten NAEP fields. Almost all the tests are multiple-choice, and, finally, the total geographic coverage falls far short of NAEP's national scope.

The college admission tests (SAT and ACT) have received much attention in recent years, primarily because of the downward trend in their scores. But their significance, except for the individual college admission candidates, has been greatly exaggerated. In the first place, only about one-quarter of the age-group take the tests, and the character of this tested group changes. This is true to such a degree that it took a special commission under Willard Wirtz two years to prepare a report on the possible significance of the SAT score decline. (For the first period of the decline, the commission found that the decline could be accounted for almost entirely by the change in the tested group.) Obviously, then, the college admission tests do not provide the kind of data that NAEP provides.

In sum, none of the other testing programs could provide the data that NAEP provides. NAEP's body of data is unique because it is national, it covers ten academic fields, and it is domain-referenced.

The Cost of NAEP

Even though NAEP's substantial achievements are unique, and its potential greater than so far realized, it is fair to ask whether it is worth its cost, currently about four million dollars a year, not counting the value of the time invested by school personnel and students.

This is a sizeable sum when looked at in isolation. But in terms of the 47 million students enrolled in elementary and secondary schools, it is less than nine cents per student! One can also compare the cost of NAEP to the estimated cost of all educational measurement — between $150 million and $200 million (MEASURING THE QUALITY OF EDUCATION by Willard Wirtz and Archie Lapointe, 1982). Or one can compare it to the total cost of elementary and

secondary education, currently over $100 billion.

Compared to other endeavors, elementary and secondary education spends a small share of its funds on assessing the product. In terms of its results, NAEP would be worth several times its current cost. In fact, it would be desirable to expand NAEP provided it was done in certain ways. (This matter is discussed further below.)

Use of NAEP Data

It is obvious but worth stating that NAEP is only as useful as the use made of it. In other words, for NAEP to move from the production of censuslike data to the improvement of educational achievement (its valid objective), others have to take the data and do something with them. What categories of people are we talking about? Journalists, national officials, state and local officials, professional education groups, independent researchers and scholars, and the public at large. I have not made an extensive survey of this use, but my impression is that the use made of NAEP data by most of these groups has been far short of the potential.

Journalists have apparently devoted little space to NAEP, at least since its early days. What few articles appear in the general press seem to deal almost entirely with the comparative data: which region did best, whether the results are better or worse than in the previous assessment, whether girls or boys did better, and so forth. This treatment exaggerates the significance of most of these comparisons and tends to ignore a very important fact: how little the results have changed over the period of the assessment (itself a collaboration of the technical soundness of the NAEP process). What journalists have failed to pay much attention to in recent years is the level of the achievement. For example, we get the information in the newspapers that the reading achievement of 13-year-olds and 17-year-olds has not changed in recent years, and that the achievement of 9-year-olds

has improved. What we do not get is information on the level of this achievement. Why not? Perhaps it is because the NAEP press releases have this same emphasis.

Looking at this question another way, the press seems to have paid much more attention, in discussing scholastic achievement, to SAT than to NAEP — despite the fact that the NAEP data is much more meaningful in assessing national educational achievement. Why is this? Several hypotheses suggest themselves. First, SAT has been around longer than NAEP, and is therefore more familiar. Second, SAT is more familiar particularly to the middle and upper classes, including the press, because almost all of them have had direct experience with SAT, whether themselves or through their children. Third, and unfortunately this suggestion does not reflect favorably on the fourth estate, those famous six digits make stories on SAT easy to write and dramatic. It takes more effort to study the NAEP data and write good copy on what they mean. It is my impression that NAEP has not done very well in facilitating this effort. For example, the latest NAEP annual report (1981) devotes only 5 of 24 pages to the achievement data itself, and well over half of those five pages ~~give~~ deals with the relatively easy, relatively meaningless comparisons (Group A did better than Group B, Group C improved in one respect over the previous assessment, and so forth). In the section on reading, there is not one non-comparative statement about how well anyone reads! Furthermore, there is no treatment of the results of assessments in five of the NAEP fields. While it is true that no new data was released in these fields during the year, this was an opportunity to summarize data from the rest of the NAEP program. After all, the last summary of NAEP data for all fields seems to be a report issued five years ago (1977)!

I have seen little evidence of use of NAEP by national officials. As for

state and local officials, there has been considerable irdirect use of NAEP data. Some 36 states have established state testing programs, and 11 of these have used NAEP methods for their own state assessment programs. Accordingly, even though NAEP itself develops no state data, it is indirectly producing a lot of state data through the state programs.

Professional education groups have used and misused NAEP data. The predictable misuse of the data has been in the nature of attempts to show that "all is right" with their particular world. A notorious example is a 1979 publication by the International Reading Association, NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS IN READING. The summary of the NAEP reading assessments therein consists entirely of a relatively inconsequential comparison between the 1970-71 and 1974-75 assessments which says nothing at all about the level of reading revealed by these two assessments. At the end of the booklet there are a number of "reflections" by reading educators, most of which are self-serving. A specialist in elementary education, for example, says that NAEP reading data provide elementary school teachers with "a pat on the back" for continuing to do a good job in teaching basic reading skills and "serve to identify various sections of the population from which children can and cannot be expected to read"! Unfortunately for all of us, many educators in recent years have been interested in identifying individuals and groups who "cannot be expected" to learn.

In contrast, a sophisicated and extremely detailed analysis of the first assessments in writing, reading, and literature was published by the National Council of Teachers of English in 1975 (NATIONAL ASSESSMENT AND THE TEACHING OF ENGLISH by John C. Mellon). This publication is an example of proper use of NAEP results: an extensive and careful analysis of both the assessment process and the educational achievements revealed. While one might quarrel with its conclusion that there is "no evidence indicating that . . . students are not learning what they ought to learn when they ought to learn it," it

recognizes that others may reach contrary conclusions. The book presents a lot of material that one can "get his teeth into".

Turning to the use of NAEP data by independent researchers and scholars, a good example is the recent report on writing made to the Ford Foundation by James D. Atwater (BETTER TESTING, BETTER WRITING, 1981). In this booklet, Mr. Atwater outlines the value of some of the NAEP techniques in assessing writing.

As for the public at large, it well may be true, as Wirtz and Lapointe claim, that only one or two percent of Americans have even heard of NAEP. Since NAEP has produced a wealth of important data over the last twelve years, it should be a matter of considerable concern that the public seems to have so little knowledge of it. Why is this, and what can be done about it? One obvious factor is the sparse media coverage that has already been referred to. Another factor is the way results are released by NAEP. There ought to be available at all times to the public and its representatives (legislators, school officials, and journalists) an up-to-date summary (perhaps no more than 50 or 60 pages) of the NAEP results to date in all ten fields of coverage. This document should be revised annually. No such publication exists. As noted above, the last summary of NAEP data for all fields seems to be a report issued five years ago!

Furthermore, while NAEP should not be in the business of interpreting its results (beyond a technical interpretation), it should be able to do more than it has been doing in the way of stimulating comment and interpretation by others. This could be done by holding conferences to discuss NAEP results and by doing more to facilitate media coverage.

Finally, the proposal by Wirtz and Lapointe that an Educational Assessment Council be established is worth serious consideration. But such a Council should be independent of government and made up entirely of

disinterested persons. What an ideal project for a foundation! Such a group, if made up of prestigious and distinguished individuals, could gain considerable public attention for the level of educational achievement in this country.

Recommendations for Expansion of NAEP

For reasons outlined above, I do not agree with Wirtz and Lapointe that NAEP is not worth continuing as is. While it is not realizing its potential, partly because of its current stripped-down character, it would be unwise to let it die, lose the continuity, and take on the added costs that would be required to start it up again at some future date. Even in its present form, it is worth the relatively modest cost.

I do believe that NAEP should be expanded because expansion along certain lines would improve its worth with reasonable increases in cost.

First, assessments in each of the ten fields should be conducted on a five-year cycle.

Second, the young-adult group (aged 26-35) should be included in all regular assessments. While this group is more expensive to survey because it is not in school, there is value in identifying the extent to which achievements are retained and the achievements that improve through further education and out-of-school experience and learning.

Third, the omission of a foreign language assessment should be remedied. Foreign language in our schools has unfortunately declined. Nevertheless, in today's world, educational achievement in modern foreign languages is important. Accordingly, I recommend that an assessment of foreign language achievement be developed for the higher two age groups (17 and 26-35). Because of the number of languages involved, this assessment would necessarily be different from the other assessments, but some meaningful program could be worked out.

Fourth, as suggested in a previous section, an annual document should be published which provides an up-to-date summary of perhaps 50 or 60 pages of the NAEP results to date in all ten (eleven if foreign languages is added) fields of coverage. While comparisons of groups and over time should be included, the emphasis should be placed on the current national level of educational achievement of the four age levels. This document, which usually would require only modest revisions each year, would be the basic NAEP report to the public and would allow legislators, school officials, journalists, and the public in general to have ready and reliable answers to the question: What are the educational achievements of our children and young people? (With background and introductory material, the report might be as long as 80 pages.)

Fifth, the complete text of more assessment exercises should be published as part of each assessment report. The exercises themselves provide concrete information which is very meaningful to laymen and professionals alike. Apparently very few are now being published. For example, in the April 1981 report, THREE NATIONAL ASSESSMENTS OF READING: CHANGES IN PERFORMANCE, 1970-80, only three exercises are shown out of 55 given to 9-year-olds, three out of 71 given to 13-year-olds, and three out of 71 given to 17-year-olds. As another recent example, in the July 1982 report, READING COMPREHENSION OF AMERICAN YOUTH: DO THEY UNDERSTAND WHAT THEY READ?, only 8% of the exercises given to 9-year-olds are shown, only 18% of the exercises given to 13-year-olds, and only 16% of the exercises given to 17-year-olds. I recommend that the full text and national results of about one-third of the exercises of each assessment be included in the published report. Naturally this would increase NAEP costs because of the need to develop additional new exercises for each cycle, but it would be well worth it.

Other Recommendations

There are other steps that can be taken to improve NAEP that do not involve significantly increased costs.

First, as mentioned in a previous section, the recent assessment reports place too much emphasis on comparisons between groups and across time. While these comparisons are interesting and easy to make, their significance has been exaggerated. More emphasis should be placed on the level of achievement by the four age-groups.

Second, as mentioned in a previous section, NAEP should do more in the way of stimulating comment and interpretation on the results by others. This can be done by holding conferences and by doing more to stimulate media coverage.

Third, as mentioned in a previous section, the proposal for some kind of Educational Assessment Council should be considered. Since I believe that any such group should be independent of government and financed by foundation funds, no additional NAEP funds would be required.

Last but not least, NAEP should not interpret its results beyond a technical interpretation. In other words, neither NAEP personnel nor a selected few non-NAEP people (who usually seem to be interested parties) should tell the public in the assessment reports or via press releases or other public statements whether the assessment results are good, bad, or indifferent, or what they mean in terms of what we have been doing educationally or what we should do. NAEP should be scrupulously neutral on these matters and should confine its activities in this area to encouraging and facilitating individuals and groups to come to their own conclusions. NAEP should, however, as a matter of public convenience, compile and disseminate complete bibliographical material on the reports, articles, and statements by others on NAEP results. In recent years, there has been a tendency for NAEP officials and NAEP reports to be apologists for the educational status quo. They should be neither apologists nor critics, but rather neutral reporters of reliable and objective information.